DOCUMENT RESUME

ED 401 626 EA 028 062

AUTHOR Mitchell, Karen J.
TITLE Reforming and Conforming: NASDC Principals Discuss School Accountability Systems.
INSTITUTION Rand Corp., Santa Monica, CA. Inst. on Education and Training.
SPONS AGENCY New American Schools Development Corp.
REPORT NO ISBN-8330-2440-X; RAND-MR-716-NASDC
PUB DATE 96
NOTE 60p.; For a related report, see ED 392 153.
AVAILABLE FROM RAND, Distribution Services, 1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138; telephone: (310) 451-7002; fax: (310) 451-6915; internet: order@rand.org
PUB TYPE Books (010) -- Reports - Research/Technical (143)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Accountability; *Administrator Attitudes; *Educational Assessment; Educational Innovation; Elementary Secondary Education; *Evaluation Criteria; Outcomes of Education; *Performance; Principals; Standardized Tests; Standards; *Student Evaluation
IDENTIFIERS New American Schools Development Corporation

ABSTRACT

The New American Schools Development Corporation (NASDC), a private nonprofit corporation, was created in 1991 to fund the development of new designs for high-performance schools. Since 1993, the designs have been implemented at over 200 schools nationwide. This book presents the findings of a formative evaluation conducted by RAND Corporation for NASDC. The study explored the elements of state and district accountability systems that facilitated and impeded innovative practice. Data were derived from visits made to 30 NASDC sites during spring 1995 and from telephone interviews conducted with 20 NASDC-school principals in summer 1995. A majority of the principals believed that standardized, multiple-choice tests failed to address the knowledge, skills, and behaviors that innovative programs seek to promote. In contrast, 85 percent of the principals in jurisdictions with performance-based accountability tests pointed to the close alignment that existed between reform goals and their states' assessments. They contended that performance-based tests model and support good classroom practice and help faculty focus on performance standards and student outcomes. The principals also identified problems associated with the reporting of school data on attendance, mobility, disciplinary referrals, course completion, graduation, promotion, and other student outcomes. It is recommended that NASDC and other reform sponsors educate and lobby at the state and district levels, to ask that accountability requirements be supplemented by explicit, congruent indicators. (Contains 16 references.) (LMI)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

RAND

E1

ED 401 626

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

J. Ryan

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

EA 028062

BEST COPY AVAILABLE

2

The research described in this report was supported by the New American Schools Development Corporation.

ISBN: 0-8330-2440-X

© Copyright 1996 RAND

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from RAND.

RAND is a nonprofit institution that helps improve public policy through research and analysis. RAND's publications do not necessarily reflect the opinions or policies of its research sponsors.

Published 1996 by RAND
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
RAND URL: http://www.rand.org/
To order RAND documents or to obtain additional
information, contact Distribution Services:
Telephone: (310) 451-7002; Fax: (310) 451-6915;
Internet: order@rand.org

REFORMING AND CONFORMING

NASDC Principals Discuss School Accountability Systems

Karen J. Mitchell

Supported by the New American Schools Development Corporation

Institute on Education and Training

4 RAND

PREFACE

The New American Schools Development Corporation (NASDC), a private nonprofit corporation, was created in 1991 as part of the America 2000 initiative to fund the development of new designs for high-performance schools. NASDC funded teams to develop and demonstrate whole-school designs to increase student performance. In fall 1993 the teams began implementing programs in 19 states and over 140 schools. By September 1995 they had negotiated partnerships in ten new jurisdictions and with 100 additional schools.

NASDC asked RAND to carry out a formative evaluation of its efforts at whole-school transformation. During spring 1995 site visits, designers, school administrators, and teachers at NASDC sites suggested several ways that state and district accountability systems might be at variance with New American Schools and other whole-school reform programs. This study sought to explore the elements of school accountability systems that aid and impede innovative practice. It also elicited school reformers' recommendations for simultaneously reforming schools and conforming to jurisdiction accountability systems.[1]

This report should interest educational policymakers at all levels of government, school administrators and teachers, education reformers, and communities concerned about improved schooling.

[1]A more extensive discussion of the reform efforts of these and other NASDC schools appears in *Lessons from New American Schools Development Corporation's Demonstration Phase* (Bodilly, 1996).

The research was supported by NASDC. The study was conducted under the auspices of RAND's Institute on Education and Training.

CONTENTS

Preface iii

Summary vii

Acknowledgments xi

Chapter One
INTRODUCTION 1
Background 1
Study Rationale and Objectives 2
Methods 3
Data Source 4
Organization of the Report 5

Chapter Two
JURISDICTION ACCOUNTABILITY SYSTEMS 7
Traditional Tests 7
Performance-Based tests 8
Other Accountability Measures 8
Score Data from Jurisdiction Assessments 8
Traditional Tests 9
Performance-Based Tests 10

Chapter Three
IMPACT OF TRADITIONAL ACCOUNTABILITY TESTS ON SCHOOL REFORM 13
Impact on Classroom Practice 13
Impact on School-Based Assessment 15
Impact on Design Implementation 16
Summary 18

Chapter Four
IMPACT OF PERFORMANCE-BASED ACCOUNTABILITY TESTS ON SCHOOL REFORM 19
Impact on Classroom Practice 19
Impact on School-Based Assessment 20
Impact on Design Implementation 22
Administration of Traditional and Performance-Based Tests 23
Summary 23

Chapter Five
IMPACT OF OTHER ACCOUNTABILITY MEASURES ON SCHOOL REFORM 25
Impact on Classroom Practice 25
Impact on Design Implementation 27
Summary 28

Chapter Six
DESIGN-BASED ASSESSMENT AND SCHOOL REFORM 29
Summary 31

Chapter Seven
RECOMMENDATIONS FOR REFORMING AND CONFORMING 33
Accountability Testing 33
Accountability Reporting 34
Recommendations to Fellow Reformers 34
Recommendations on Accountability Requirements . . . 35
Recommendations on Reforming 40
Recommendations to NASDC 43
Conclusions 46

References 49

SUMMARY

BACKGROUND

The New American Schools Development Corporation (NASDC) has funded the development of seven comprehensive designs for high-performance schools. In June 1993 NASDC began implementing programs in 19 states and over 140 schools. The designers and their site partners spent two years piloting and refining the reforms. In the spring of 1995, NASDC negotiated partnerships in ten new jurisdictions and with 100 additional schools. These schools began their work with NASDC in fall 1995. RAND has examined the development, initiation, and implementation of these whole-school reforms. Many of their elements are common to programs sponsored by other reform networks and by innovators in charter schools, state district program offices, and elsewhere.

STUDY RATIONALE AND OBJECTIVES

RAND's spring 1995 site visits with designers, school administrators, and teachers at 30 of the original NASDC sites suggested several ways that state and district accountability systems might be at variance with New American Schools and other whole-school reform programs. Reformers said that system-sponsored standardized, multiple-choice tests were misaligned with teaching and learning as conceived in their programs. Some asserted their state or district accountability system narrowed the focus of community discourse from their broad, comprehensive reform goals to things assessed by the system and profiled in accountability reports.

METHOD

This study sought to more fully explicate and explore the elements of school accountability systems that aid and impede innovative practice. In the summer of 1995, follow-up telephone interviews (~45 minutes) were conducted with 20 of the original NASDC principals. Principals were asked to reflect on the impact of their accountability systems on classroom practice, student assessment, design implementation, and school restructuring. They were asked to offer recommendations to fellow reformers and state and district sponsors for simultaneously reforming schools and conforming to jurisdiction accountability systems. The study was designed to help administrators in present and future innovating schools anticipate and plan for the possible discord between reform aims and their jurisdictions' accountability requirements.

FINDINGS

NASDC principals made a number of arguments about the impact of traditional accountability tests on school reform. The large majority said that standardized, multiple-choice tests are misaligned with the classroom practices of reforming schools and that traditional tests do not address the knowledge, skills, and behaviors innovative programs seek to promote. Over half said that state and district tests drive the educational program to focus on basic skills and use traditional strategies. Several asserted that preparation for traditional accountability tests takes time away from design-prescribed instruction, meaningful learning, and design implementation. Eighty-five percent of respondents said their teachers "stepped out of the design" to prepare students for mandated tests by reviewing skills and content likely to be on the test, practicing test-taking skills, and working with practice tests and test preparation materials.

Some respondents also said that traditional tests provide teachers who are reluctant to reform another excuse to hold onto old ways. Further, they argued that systems designating "accountability" grades for testing pose problems for multi-age/grade designs; respondents explained that third graders exposed to the fourth grade curriculum still have to take the third grade test; so do third graders working on second grade units.

In contrast to these assertions, 85 percent of the respondents in jurisdictions with performance-based accountability tests talked about the close alignment between reform goals and their states' assessments. They contend that performance-based tests model and support good classroom practice and help faculty focus on performance standards and student outcomes. Principals credited authentic accountability measures with helping reform-minded teachers internalize standards, become accustomed to performance assessment, develop open-response tasks and rubrics, and eschew classroom-level, multiple-choice testing.

In addition to test scores, many accountability systems report school data on attendance, mobility, disciplinary referrals/suspensions, promotion, graduation, advanced course completion, and other student outcomes. Principals said that some of these measures pose problems for reforming schools. They made several points: (1) Judgments based on grade-to-grade promotion rates are inconsistent with multi-age, multi-grade designs. (2) Reporting numbers of students earning specified numbers of credit hours is problematic because it is hard to assign credits for large teaching blocks; also, it is hard to give disciplinary credits for multi-disciplinary activity. (3) Credit hours in some jurisdictions are linked to numbers of hours in the classroom, and some reform programs' community-based components complicate "in-class" indicators. (4) Some accountability systems have targets for the numbers of hours teachers and students should be "in contact"; contact-hour targets make it difficult to arrange schedules allowing for joint teacher planning time. (5) The need to provide high school students with credits, competitive standardized test scores, and interpretable transcripts so they can get into good colleges is an impediment to innovative practice. (6) Reporting teacher attendance rates among accountability data is a problem for schools encouraging participation in reform-relevant teacher professional development.

CONCLUSIONS

Study participants talked about the elements of school accountability systems that aid and impede innovative practice. In the same way that school-based reformers are looking for ways to educate proximate stakeholders—parents, community members, and business

partners—about their students' accomplishments and the significance of jurisdiction accountability requirements, NASDC and other reform sponsors should educate and lobby at the state and district level. Principals asked that ill-fitting accountability requirements be supplemented by indicators consonant with reform.

To date, NASDC has only generally addressed the barriers to reform posed by its partner jurisdictions' accountability systems. NASDC and others propelling reform should work with jurisdiction leaders to develop methods for fairly gauging the impact of reformers' work on students, teachers, parents, and the larger community in which these schools operate.

ACKNOWLEDGMENTS

I would like to thank the eloquent principals whose efforts and ideas are reflected in this report. I appreciate the generosity of their intellect and time. I am also grateful to the New American Schools Development Corporation and the Ford Foundation for their support of this study. RAND colleagues Thomas Glennan, Susan Bodilly, and Susan Purnell offered many helpful suggestions throughout. They and the report's reviewers, Lawrence Hanser and Paul Hill, made important contributions to the improvement of the report. Finally, thanks go to Susan Spindel for her excellent preparation of the document.

This report is dedicated to the memory of Elsbeth Kehl, who brought intelligence, dedication, and enthusiasm to her work at NASDC and with reforming schools.

Chapter One

INTRODUCTION

BACKGROUND

The New American Schools Development Corporation (NASDC) has funded the development of seven comprehensive designs for high-performance schools. In their 1991 solicitation of reform models, NASDC specified that designs were to integrate all elements of a school's life (NASDC, 1991). They were to address whole schools—not just a single grade or program within a school. Designs were to be benchmarked against demanding goals and achievement standards. The designs were to be for all students, not only those most likely to succeed. The seven designs that resulted represent differing philosophies and varied, research-based education practices (NASDC, 1994). Unifying the diverse designs, however, are several principles, including:

- High academic standards,
- Curricular and instructional strategies that include thematic, project-based, and interactive learning,
- Performance-based assessment systems,
- Continuous professional development for teachers and staff,
- Service to and support from parents and the community,
- School autonomy and decentralized governance structures, and
- Integrated use of technology to enhance student, teacher, and school performance.

These elements simultaneously describe the school as an organization and schooling as a process. They are evident in other whole-school reforms, including those of the Coalition of Essential Schools, the Accelerated Schools Project, and the Comer School Development Program. These elements describe the work of innovators in charter and other distinctive schools; they are common to reforms outlined by district program offices, policy institutes, teachers' unions, and others.

In June 1993 NASDC began implementing the seven designs in 19 states and more than 140 schools. The design teams and their site partners spent two years piloting and refining the reforms. NASDC asked RAND to examine the development, initiation, and implementation of the designs. RAND's document reviews, discussions, interviews, and focus groups provided information about the progress and challenges of the programs (Bodilly, 1996). During school years 1993/94 and 1994/95, RAND observed that the teams and schools made rapid progress. Progress was hastened, analysts proposed, because the designs put forth broad visions of reform and school change; because transformations involved entire schools, rather than smaller programs or design features; and because substantial technical assistance and professional development were provided.

In spring 1995, NASDC and the teams entered a new phase of their work. They negotiated partnerships in ten new jurisdictions and with 100 additional schools. These schools began implementing designs in fall 1995 (NASDC, 1995).

STUDY RATIONALE AND OBJECTIVES

Spring 1995 site visits with designers, school administrators, and teachers at 30 of the original NASDC sites suggested several ways that state and district accountability systems might be at variance with New American Schools and other whole-school reforms. Reformers said that system-sponsored standardized tests were misaligned with teaching and learning as conceived by their programs. They said that multiple-choice tests were not synchronized with their schools' classroom practices and did not reflect students' accomplishments. Some asserted their state or district accountability system narrowed the focus of community discourse from their broad, comprehensive

reform goals to things indexed by the system and profiled in accountability reports.

This study seeks to more fully explicate and explore the elements of school accountability systems that aid and impede innovative practice. It examines reformers' efforts and the requirements of ten typical jurisdiction-sponsored accountability programs. The study is designed to help administrators in present and future reforming schools anticipate and plan for the possible discord between reform aims and their jurisdictions' accountability systems. Results may help participants in NASDC and other innovative programs establish reasoned expectations for school progress and outcomes. The views and recommendations of these initial NASDC reformers may also prompt NASDC and other reform leaders to identify and advocate for system supports that enable initial and sustained school transformation.

METHODS

In summer 1995, telephone interviews (~45 minutes) were conducted with a subset of the site visit principals. Protocols were designed to elicit more detailed data than had been possible in spring when a broad range of topics was addressed. Semi-structured interviews began with questions about the traditional and performance-based tests mandated by the district and state; alignment between mandated assessments and instruction as prescribed by the design; stakes attached to the school's performance on mandated assessments; and publication of district- or school-level profiles by the jurisdiction. Principals were asked to discuss four broad issues:

- The impact of state and district accountability systems on classroom practice,
- The impact of state and district accountability systems on student assessment as prescribed by the design,
- The effect of state and district accountability systems on design implementation and reform efforts generally, and
- Their recommendations for simultaneously reforming schools and conforming to state and district accountability systems.

Questions were asked in an open-ended format to elicit broad and complete statements of the impact of school accountability systems on reform. The methods were exploratory and relied on previously collected school information and established relationships with site visit principals. The study design does not support generalizations about the numbers or types of NASDC or other transforming schools aided by or struggling with particular accountability requirements. A follow-on study with more structured questioning and a larger and non-overlapping respondent pool is needed to support stronger statements about accountability and school reform.

Because principals' discussions were largely extemporaneous (not responses to specific prompts about the features of accountability systems), the number of study principals who support described views potentially underestimates the number holding that view. Unless specifically noted in the findings, the proportion of respondents expressing a particular view are those electing to express the viewpoint, not the proportion of study principals concurring.

All interviews were conducted by the author, who used shorthand to record responses. Interview records were then transcribed, resulting in 190 pages of interview text. Coding categories were developed for each question and responses were double-coded.

DATA SOURCE

In fall 1993 RAND designed a sample for and began comparative case work with 30 NASDC schools. For each design two units of observation were selected—schools or feeder patterns or districts—as suggested by the designs' intervention points. That is, some designers worked with individual schools in differing districts; for each of these designs, two individual schools were included in the sample. Some teams entered through feeder patterns that included the elementary and middle schools feeding into a high school. Other designers negotiated partnerships at the district level. In these cases, feeder patterns or districts were sampled as appropriate. The case study sample included urban, rural, elementary, middle, secondary, low poverty, and high poverty schools. It serves as the base for the present study.

Five of the original 30 schools were omitted from the interview pool—one because it was a kindergarten-only site, another because it began implementation in the second, rather than first, implementation year, and three because they were headed by first-year principals in 1994/95. Principals at 20 of 25 schools implementing NASDC's seven designs completed interviews (80 percent).

The principals and their staff began implementing designs in school year 1993/94; some also helped develop and refine design concepts and products that year and the following year. The telephone interviews were conducted after the second year of NASDC participation. At the time of the interviews, all principals had received spring 1994 accountability data for their schools. Most had not obtained 1994/1995 accountability information. Principals' statements about school transformation and accountability reflect their experience during both reform years; their statements about school accountability results largely reflect data from their first NASDC year.

To help preserve respondents' anonymity, references to individual designs are omitted in the report; designs are referred to generically as NASDC designs. Female pronouns are used throughout. To the fullest extent possible, findings are presented in the language of the 20 NASDC principals.

ORGANIZATION OF THE REPORT

The elements of the state and district accountability systems under which these sites operate are described in Chapter Two, along with the principals' assessments of the impact of traditional accountability tests on school reform. The impact of the tests on classroom practice, school-based assessment, and design implementation is discussed in Chapter Three. Chapter Four presents principals' assessments of the impact of performance-based accountability tests on school transformation. Chapter Five outlines the impact of other accountability measures on classroom practice and reform. Chapter Six treats the role of design-based assessments in school reform. In Chapter Seven, the 20 NASDC principals offer recommendations to like-minded educators and reform sponsors for simultaneously reforming and conforming to school accountability systems.

Chapter Two

JURISDICTION ACCOUNTABILITY SYSTEMS

The NASDC designs exist in jurisdictions that administer varying combinations of norm- and criterion-referenced tests, on-demand performance tests, and student portfolios. In many states and districts, other student data, including attendance rates, promotion rates, and graduation rates, are also evaluated by accountability systems. In most jurisdictions, test and other accountability data are reported in district and/or school profiles.

TRADITIONAL TESTS

Eighteen of the 20 study schools (90 percent) are in jurisdictions that include standardized, multiple-choice tests among their accountability measures. In these jurisdictions, data from traditional tests are reported alone or in combination with scores from performance-based tests. The standardized, multiple-choice tests[1] administered in these jurisdictions include the Abbreviated Stanford Achievement Test (ASAT), the Iowa Test of Basic Skills (ITBS), the Indiana State Test of Educational Performance (ISTEP), the Massachusetts Educational Assessment Program (MEAP), the Maryland Functional Test (MFT), the Stanford Reading Test (SRT), and the California Test of Basic Skills (CTBS). Additional standardized tests are given in some schools through Chapter 1 programs. Two of the ten study districts also require administration of criterion-referenced tests.

[1] In this report a number of labels for traditional, multiple-choice tests are used interchangeably: standardized tests, fixed-response tests, traditional tests, and conventional measures.

PERFORMANCE-BASED TESTS

A dozen schools in this group (60 percent) administered state-sponsored performance-based tests in 1993/94. Performance-based tests require students to create answers or products to demonstrate their knowledge and skills. Rather than choose from presented options as for traditional, multiple-choice tests, students must create or construct responses to performance-based tests.

Schools in the study group administered the California Learning Assessment System (CLAS), the Kentucky Instructional Reform Information System (KIRIS), the Maine Educational Assessment (MEA), and the Maryland School Performance Assessment Program (MSPAP) in spring 1994. All but two of the twelve schools administering performance-based tests in 1993/1994 also gave conventional tests. The CLAS tests were canceled by the California legislature in 1994 and were not administered in spring 1995. Performance-based tests will be referred to in this report as performance assessments, authentic assessments, and performance batteries.

OTHER ACCOUNTABILITY MEASURES

In addition to test scores, many accountability systems record data on the following: student attendance, mobility, disciplinary referrals/suspensions, promotion, inclusion, graduation, number of first graders with kindergarten experience, number of students performing at grade level, advanced course completion rates, and number of students with two or fewer failing grades. Some systems also track community partnerships, volunteerism, and parent involvement in the school. The jurisdictions of three-quarters of the study sites produce or mandate school profiles or school report cards; along with test information, almost all of the profiles include data on other accountability measures.

SCORE DATA FROM JURISDICTION ASSESSMENTS

Principals were asked about their schools' assessment data for school years 1993/94 and 1994/95. The school summaries provide a backdrop for principals' statements about the accountability measures

that aid or impede their efforts. Again, at the time of the interview, all had spring 1994 data; few had spring 1995 results.

Traditional Tests

On the traditional assessments, eleven principals said they received good or improved test data for 1993/94 (61 percent); seven reported mixed or low scores (39 percent). Principals were asked whether the score data they received corresponded to their own observations about school performance. These are typical responses:

> We got back data for 1993/94 and 1994/95 and they were as expected. We have always done pretty well. We were doing well on ISTEP before NASDC. For 1992/93, we were a four-star school; we were in the top 25 percent of schools based on scores and attendance.

And

> We got ITBS results in 1994 and 1995. They were well above average in both years. Because we worked on it. We need to make the test scores improve because . . . the district wants the scores to improve for reforming schools.

And finally,

> We had heard from others, and I believed, that if you teach kids to think critically and analyze, they will do well on any kind of test. But this did not pan out. Even though there was a lot of learning going on. The students were engaged, discipline problems dropped by two-thirds, attendance rose, but their test scores went down. We have a new superintendent. She looked at last year's scores and compared them to this year's. She then read out the difference scores at an administrative council meeting. She read them out from top to bottom (high positive gain to high negative change). My stomach churned as I listened and listened and didn't hear my school. My stomach is still churning.

Like the latter respondent, four of the principals with disappointing score data said they believed at the outset of their work with NASDC that students doing rigorous work, thinking critically, and analyzing

well would be prepared for any test—traditional or otherwise. Two years into the reform, they said they question that view.

Nine of the principals (50 percent) said local newspapers reported their schools' multiple-choice test data in 1994 without accompanying information about student populations or the schools' vision; score summaries were published in such a way that they became the 'descriptor of the year." Among other remarks, a principal made this one:

> The scores are published in the paper. Pandora's box is opened when the media gets hold of them. They rank you in relation to the other elementary schools. This is a bone of contention if you are not in the middle or above . . . It is such a joke because we can't show growth in relation to reform goals and earlier performance.

Interviewees lamented this practice; they said it narrowed the focus of community attention from their programs' aims and reform progress to performance on tests that misrepresent students' achievements.

Score data on traditional tests for 1994/95 had been received by four principals prior to the interviews; two schools obtained positive results and two obtained discouraging summaries.

Performance-Based Tests

On the performance batteries, results were reversed. Seven of the principals (58 percent) reported mixed or low performance assessment results; the remainder saw good or improved results (42 percent). These data are not surprising given the high, forward-looking standards used to score the four states' performance tests. Conspiring for disappointing results was the fields' relative inexperience with open-response accountability tests in spring 1994, as well as the nascent state of the designs and their implementation.

Principals talked about their schools' responses to the data. About positive reports interviewees said,

> On CLAS '94 the kids did good and this makes sense because it is the way we are teaching . . . Last year on ASAT . . . results were mixed . . . In 1994 we expected our kids to do better on CLAS

> because it is more problem-solving oriented. ASAT is not the way we are teaching and people understand that. The district knows we are trying to reform and they are patient.

And

> In 1994/95 the scores were high and showed us that the strategies we applied were working. When we got the scores back, we felt good. We wanted the kids to achieve at high levels.

In contrast, several principals told of vexing results.

> In 1994 we got CLAS results and it was disappointing for all of us. We did a little better than the state and district in a couple of areas, but our math was poor . . . It wasn't what we expected. We couldn't come up with any explanations for it. We couldn't make heads or tails of it.

Similarly,

> We got back KIRIS and Basic Skills Test results. In 1993/94 our writing (score) was way below what I thought it would be. I was surprised. These low scores had a significant impact. The writing was shocking. I met with the SBTM council (site-based management team) and we decided we needed to hire a writing specialist.

Of principals in jurisdictions where accountability systems include both performance-based and traditional tests, some reported non-converging data across tests—that is, the two sets of test data painted different pictures of school performance. More reported converging results on traditional and authentic tests; however, the majority of principals with converging data reported low scores on both types of tests. Two principals had received spring 1995 performance assessment data before the phone interviews. Both received encouraging reports.

Chapter Three

IMPACT OF TRADITIONAL ACCOUNTABILITY TESTS ON SCHOOL REFORM

IMPACT ON CLASSROOM PRACTICE

NASDC principals were asked about the impact of traditional, multiple-choice accountability tests on classroom practice as prescribed by their designs. Fourteen principals in schools with system-sponsored standardized, multiple-choice tests said these tests are misaligned with their programs (78 percent). They said that standardized tests are not synchronized with the classroom practices of reforming schools; traditional tests do not address the knowledge, skills, and behaviors innovative programs seek to promote. One interviewee remarked,

> The ASAT doesn't look at problem solving or collaborative work. It is about remembering detail . . . ASAT is more basic-skills oriented and we are doing bigger things than that with our design. We are trying to be good thinkers.

Similarly,

> We have gone from a factory approach to a child-centered model of learning. My notion is that tests are what they are. They are not indicators of students' abilities. They are indicators of what kids can do on isolated tasks in a 40-minute period, on isolated words and on short paragraphs.

And finally,

> We have built a better mouse trap. Our program matches better to IPAS (Indiana's proposed performance-based test) than ISTEP. Students can't show all their skills on multiple-choice tests. We have outgrown ISTEP.

Ten of the principals (56 percent) said their state and district tests drive the school's educational program. Principals explained that accountability tests prompted faculty to address specific content, attend to given skills, and apply particular instructional strategies. In describing the influence of traditional batteries on curriculum, principals most often talked about re-inserting basic skills into the curriculum; one of the principals explained, "You have to sprinkle lower-order sugar in with the pepper of higher-order thinking stuff." Another declared,

> Next year we want to maintain some of the design elements but will make sure we hit the essential skills. We have listed the skills for teachers, and the teachers will check off which ones they hit as they are teaching units. The skills will be integrated into the curriculum; they will be basic skills though. They're lower level than higher level.

Five of the principals (28 percent) additionally argued that accountability requirements take time away from important work. They said that preparation for traditional accountability tests takes time away from design-prescribed instruction and meaningful learning. One principal remarked,

> It is a problem for reformers. We have spent quite a lot of time doing test preparation and using materials from scholastic publishers—which are totally foreign to what we are doing. We don't want our kids to bottom out because they don't think that way. The two (reform and accountability) don't mix at all . . . it takes time away from meaningful learning for kids and meaningful instructional tasks for teachers. It takes a week for the test and 30 to 45 minutes per day in the six weeks preceding the tests . . . It is a real paradox.

And

> The kids are actively involved in group learning and we had to transfer that to a sit-down test where they couldn't ask for help from anyone—from students or teachers.

In fact, in response to specific questioning, 85 percent of all respondents said their teachers "stepped out of the design" to prepare students for accountability testing. These teachers and their students, principals explained, reviewed skills or content likely to be on the test, practiced test-taking skills, and worked with practice tests and other test preparation materials. This practice wasn't peculiar to preparation for fixed-response tests; principals reported directed preparation for performance-based tests as well. In fact, last year one of the study schools received district funding for an MSPAP support teacher—a faculty member dedicated to improving the school's showing on the state performance test.

In contrast to these positions, three principals (17 percent) downplayed the effects of conventional tests on innovative practice. They countered their colleagues' arguments, saying that traditional accountability tests have little effect on transforming schools' classroom practices. These principals said teachers thought of the accountability assessment as a "pill they had to take." One respondent explained,

> We don't spend a lot of time trying to align the curriculum to ASAT . . . We pay attention, but we don't spend a lot of time because it doesn't reflect achievement.

IMPACT ON SCHOOL-BASED ASSESSMENT

Reformers also were asked about the impact of traditional accountability tests on student assessment as prescribed by the design. On this point principals were unanimous; they asserted that traditional tests have little impact on their own testing practices. One principal spoke for the group when she said,

> We report the (standardized test) data out to staff so teachers are aware of areas where the kids are not doing well, but it is not the guiding force. We still use CLAS prompts in the classroom. We put our passion and energy into authentic assessment.

Another explained, "We do it and get it over with. My ITBS stuff is in a big heap; I did it and set it aside."

IMPACT ON DESIGN IMPLEMENTATION

Interviewees talked about the impact of jurisdiction-sponsored standardized, multiple-choice tests on school restructuring. Seven principals (39 percent)—most of whom reported positive test results for last year—said their jurisdictions' accountability systems didn't affect restructuring in 1994/95. One principal stated,

> I knew I would have to deal with it and I didn't let it seep into other things. We did have to try to motivate the kids because they could not see the relevance and it flies in the face of what we are trying to do.

Five respondents (28 percent) lamented the time lost on test preparation and administration; principals said that time diverted from reform to test preparation and administration is a drain on program implementation. One respondent said,

> It has been a deterrent. Those (conventional testing) systems take time and we get data but the time is not well spent; students can't demonstrate well what they can do. It has hindered us because of the time spent away from the design—in preparation and administration.

A couple of principals further charged that traditional testing programs give teachers reluctant to reform another excuse to hold onto old ways. One explained, it "gives another reason to avoid the design."

Respondents also spoke generally about the deleterious effects of accountability testing on school transformation. A few principals talked about using laudable test data to "buy" their faculty the flexibility necessary for innovation. They said that scoring well on accountability tests is taken by parents, community members, and others as collateral for the "luxury of reform." One principal explained,

> If test scores are low and parents are not pro-NASDC, they jump on the bad scores. We have to do well on the dreaded standardized tests if we want to have the luxury of doing the kind of work we want—multi-age, exhibitions, portfolios.

And

> Good scores help you in getting the curriculum changes you want through the parents and community. Our parents are pretty conservative; their motto is "do unto my kids as you did unto me."

In this vein, a couple of respondents talked about disappointing results and their impact on system support; one professed,

> The scores have caused us anguish and professional dissonance. We worked so hard last summer and during the school year; the teachers worked non-stop . . . Even though there was a lot of learning going on . . . the test scores went down . . . The school is at risk of being on probation in the district . . . Maybe the low scores mean we've done a good job of implementing the design—but that's hollow thanks.

Two principals reported requests by their districts to administer additional tests as a check on reform progress; at one of these schools, the principal declared,

> The district is just requiring this of me because of NASDC. The rest of the district only did the test at specific grade levels. The data are going to kill us because the test is so content-oriented that I fully expect it to be a problem . . . I will get called in and asked to improve them. I will try to explain them away and maybe if we can hold our own, that will be enough. People today are so standardized-testing oriented.

Two other general arguments were made about accountability testing—traditional or otherwise. First, it was argued that systems designating "accountability" grades for testing pose problems for multi-age/grade designs. Fourth graders exposed to the third grade curriculum still have to take the fourth grade test, they said; so do third graders working on fourth grade units. One principal explained,

> Next year we will move away from ability-level instruction (multi-age); this goes against the design, but when you have a fourth grader who is doing work with third graders . . . they are still accountable for the fourth grade curriculum . . . so next year we will keep the students on grade level for language arts and math, but we will do cross-aging in everything else.

Second, principals pointed out that most accountability tests reinforce disciplinary divisions; available tests are inconsistent with multi-disciplinary units and student work. This argument applies almost without regard to test format (traditional or authentic). Unlike other accountability tests, the Maryland School Performance Assessment Program has tasks addressing multiple disciplines.

Finally, when asked about the usefulness of standardized test data for making inferences about school improvement, five principals (28 percent) judged standardized test data very useful, nine said they are somewhat useful (50 percent), and four described them as not useful (22 percent). Several said the scores have utility as one among a number of data points about school performance.

SUMMARY

To summarize, principals said that traditional accountability tests drive their programs in directions inconsistent with reform; they argued that conventional tests detract from teaching and learning as prescribed by their designs and divert resources from program implementation. Interviewees contend that standardized, multiple-choice accountability tests are a barrier to school restructuring.

Chapter Four

IMPACT OF PERFORMANCE-BASED ACCOUNTABILITY TESTS ON SCHOOL REFORM

Ten interviewees from the four states administering performance tests in 1993/94 talked about the close alignment between reform goals and their states' assessments (85 percent). Of the authentic tests, NASDC principals said,

> Process-wise, there is a nice match—cooperative learning, partnering—that's built into MSPAP. Also kids are allowed to arrive at answers in a variety of ways. MSPAP gives credit for putting appropriate processes into play if the ultimate answer is not correct. Our design corresponds to the processes. That part is good.

And

> We are developmentally-based. The MEAs truly test the way that we teach. We are very performance based . . . It assesses how we teach. We have spent more time preparing our students for the test because it does test what we teach. We look to see if we are missing anything.

Interviewees also talked about the impact of system-sponsored performance tests on classroom practice, student assessment, and restructuring.

IMPACT ON CLASSROOM PRACTICE

In describing the impact of performance-based tests on curriculum and instruction in reforming schools, principals declared that performance-based accountability tests model and support good class-

room practice. One principal explained, "Good practice is good practice. There is a blending of district-prescribed good practice and NASDC has added to it." Another asserted,

> It is making teachers step back and take a second look at teaching practices . . . It helps them push kids. It is making them think about curriculum and where they want kids to go.

And

> It has changed the way we think about the learning process. We have teachers using thematic units, hands-on activities, the constructivist classroom.

Finally,

> MSPAP expectations go with the vision of NASDC. Teachers live in compartments and they are used to seeing program one and program two and program three, and many times there is no connectedness. But when we looked at MSPAP and NASDC, they were the same. We weren't doing MSPAP or NASDC; we were just doing good teaching.

Twenty-five percent of the principals said that performance-based accountability tests helped their teachers focus on performance standards and student outcomes. One administrator explained,

> It (the performance test) has had a major impact. I don't know that the teachers ever really cared before—I don't mean that exactly—but now they see the state standards and what will be tested. They used to just use the textbooks before, and go through them page by page.

IMPACT ON SCHOOL-BASED ASSESSMENT

On the impact of authentic accountability tests on their schools' assessment practices, one principal made the following statement,

> MSPAP certainly has had an impact. It was a nice meeting of the minds with NASDC. They are both pushing us in the same direc-

> tion. It is fortunate for us; it is helping to take us where we want to go.

Only positive statements were made about authentic accountability tests and school-based assessment. Respondents stated that authentic accountability tests help reform-minded teachers:

- Internalize the standards,
- Get used to performance assessment,
- Develop open-response tasks and rubrics, and
- Eschew classroom-level, multiple-choice testing.

For example, principals said,

> All students create portfolios and teachers are scoring them using the state rubrics . . . We try to focus sixth and seventh grade on the eighth grade year; that is the important year, the accountability year. The system is helping us move toward a standards-based system.

And

> Within the classroom, teachers now are used to more performance-based assessments. In social studies, for example, they do performance components . . . Then they take a performance-based test in which they have to do research on a character.

Finally,

> The teachers created a CLAS-type assessment this year—along with rubrics. They gave all students, age four to twelve, one writing prompt. It was about a memory. Directions went home to the parents to help the student think about a memory and they (the students) wrote about it at school. The teachers created a 15-point score scale and picked anchor papers. The rubric was applied across grade levels.

An interviewee acknowledged her teachers use "fewer multiple-choice tests now. We'll do more and more with portfolios next year."

IMPACT ON DESIGN IMPLEMENTATION

Respondents also talked about the impact of performance-based accountability tests on design implementation. Two positive effects were described. First, principals said that reform is aided by state- or district-prescribed good practice that reinforces design tenets. Four respondents (33 percent) said the jurisdictions' prescriptions mirrored and reinforced those of NASDC; an interviewee explained, "It (the test) supported design implementation. It was a lot easier to justify things when I had two things going for me." A second principal stated,

> Very clearly, the system has helped us. We feel so fortunate. In every meeting I attend, I learn that Kentucky is so far ahead of everyone. We are already doing stuff NASDC is talking about.

Second, a few principals talked about the information value of authentic assessment data and about the credibility that positive scores accorded reforming schools. They said that favorable data from system-sponsored performance assessments inform and give credence to the school's efforts:

> The state accountability system helped because we have real data—not only perceptions that kids liked it (the instructional program). Because it comes from a higher being, it is taken with more seriousness than if we said it ourselves. There always has to be some monitoring force above for people to recognize that it is important.

And

> My role is to work with the teachers and with strategies to promote an outcomes-based culture. The system has strongly moved them to a standards-driven curriculum. It gives them feedback about how well they are doing.

When asked whether data from performance-based accountability tests are useful in making inferences about school improvement, nine of the respondents (75 percent) judged them to be very useful; three (25 percent) asserted the data are somewhat useful in thinking about school improvement; none denounced their use. These results

compare well to those for traditional tests where corresponding percentages were 30, 50, and 20, respectively.

ADMINISTRATION OF TRADITIONAL AND PERFORMANCE-BASED TESTS

As earlier noted, all but two of the twelve schools administering performance-based tests in spring 1994 also gave conventional tests. Several described the tensions raised by administration of dissimilar batteries; one respondent proclaimed, "We are operating in a schizophrenic mode." Of the Maryland accountability system, for instance, a second principal said,

> Teachers have trouble seeing MSPAP and the Maryland Functional Test as one because they are so different. Teachers have a hard time seeing how they work together. You can't get to broad generalizations if you only drill on division facts. It causes conflict for teachers because they are teaching to two different tests.

And

> Once I get them thinking the other way, I have a hard time telling them to do drill and practice. That is where the conflict comes . . . It is the bridge they still have to figure out. The math people are the ones having the hardest time with this because they are so content-focused.

SUMMARY

NASDC principals in jurisdictions with performance-based accountability tests contend performance assessments are beneficial to reform; they said that authentic assessments model good instruction, draw attention to standards and outcomes, and provide data for school improvement. Principals said that performance-based accountability tests benefit school reform.

Chapter Five

IMPACT OF OTHER ACCOUNTABILITY MEASURES ON SCHOOL REFORM

In addition to test information, many accountability systems compile data on student attendance, mobility, disciplinary referrals/suspensions, promotion, inclusion, graduation, number of first graders with kindergarten experience, number of students performing at grade level, advanced course completion rates, and number of students with two or fewer failing grades. Some systems also include data on community partnerships, volunteerism, and parent involvement in the school. Fifteen of the study sites (75 percent) operate in jurisdictions that produce or mandate school profiles. Along with test information, almost all of the profiles display data on other accountability measures. Principals talked about the impact of these noncognitive accountability indicators on classroom practice and on school transformation.

IMPACT ON CLASSROOM PRACTICE

Several principals talked about accountability measures that pose problems for transforming schools. They made five points: four are given here; a fifth is treated separately.

- Judgments based on grade-to-grade promotion rates are inconsistent with multi-age, multi-grade designs.
- Reporting the number of students earning a specified number of credit hours is problematic because it is hard to assign credits for large teaching blocks; also, it is hard to give disciplinary credits for multidisciplinary activity.

- Credit hours in some jurisdictions are linked to number of hours in the classroom, and some reform programs' community-based components complicate "in-class" indicators.
- Some accountability systems have targets for the number of hours teachers and students should be "in contact"; contact-hour targets make it difficult to arrange schedules allowing for joint teacher planning time.

By way of illustration, one high school principal explained,

> School credits are a challenge for reforming schools. In most high schools, you have to have 40 credits to graduate. There are credit requirements by discipline too. In a non-traditional program, it is hard to figure out how to assign credits because of the longer teaching blocks and the interdisciplinary curriculum. If you're doing the *faces* expedition for two hours a day, you have to give half credit for art, and a whole credit for something else, and another half for something else.

Also, NASDC principals said,

> The requirement for additional contact time messes up our scheduling . . . By increasing contact time, we lose joint planning time and that jeopardized expeditions because teachers can't plan well or they have to plan alone . . . The state requires 200 minutes of time per week for the standard carnegie unit . . . We have 200 this year and will get to 220 next year to assuage the school board.

Finally, college placement is seen as an important accountability issue by most high school communities. The need to provide high school students with credits, competitive standardized test scores, and interpretable transcripts so they can get into good colleges is an impediment to innovative practice. Principals bemoaned the fact that, at least thus far, they have been unable to effectively convey to admissions officials their programs' unique characteristics and rigor. Further, at no point in a student's educational program is exemplary performance on standardized tests more important than at entry to undergraduate school. One principal explained,

> We are trying to educate the admissions people on what we are doing and to come around to our way of thinking . . . U.S. History and

> Algebra, for example, don't show up because we have an integrated curriculum. We have to show what relates to what.

IMPACT ON DESIGN IMPLEMENTATION

Two arguments about the impact of other accountability measures on design implementation were made. First, reporting teacher attendance rates among accountability data is a problem for schools encouraging participation in reform-relevant teacher professional development. The faculty, one principal said, "have a lot of expertise to develop through lots of staff development."

And second, the student incentives relied upon by schools to improve school accountability data divert energy and resources from reform. Respondents talked about the incentives relied upon to improve school accountability data. Last year, one study school gave popcorn parties for students with 100 percent attendance, another provided pizza for students who showed up for testing, a third looked for and picked up students who missed the bus. These actions are not peculiar to transforming schools; however, they point to the possible diversion of resources from reform. One study principal described her faculty's efforts to increase attendance and promotion rates:

> We will never be rated high on MSPAP because of our poverty level, language status, and mobility rates. We will only be able to meet state standards on attendance and promotion. So we really work hard on attendance. The kids like to be here, we know that, but some of them have to stay home to baby-sit sick brothers and sisters so their parents can go to work, some have to translate during the day for their parents, and some are sick themselves and there's no money for doctors. This year we have been over 94 percent every month. We think we might get 96 percent for the year . . . but it takes a lot of planned effort on the part of a lot of people. We have home visits. We look for missing kids . . . Promotion rates are a hoot. You have data that show retention is not good, but this is a philosophical dilemma. Do you keep them back or pass them all? . . . the school might be able to get a star on that data-based area. I calculate how many kids I can retain and still get a star and then I don't go over that number.

SUMMARY

Principals' descriptions of the barriers to reform raised by noncognitive accountability measures were less impassioned than their statements about accountability tests. Nonetheless, their messages were clear: a number of these measures misrepresent the performance of transforming schools.

Chapter Six

DESIGN-BASED ASSESSMENT AND SCHOOL REFORM

By way of context and—in many cases—contrast, principals described the use of design-based assessments in their schools. Almost all talked about their students' work with portfolios; additionally, principals talked about logs and journals, ability or growth records, exhibitions and projects, open-response assessments (including writing prompts), and self-assessments and check-sheets. Several principals pointed to parent/teacher conferencing as an important element of design-based assessment.

Sixteen respondents (80 percent) said design-based assessments provide information very useful to school improvement planning; the remainder said they are somewhat useful for this purpose.

Principals said,

> We saw how powerful these kinds of assessments are rather than paper-and-pencil tests. It confirmed what some suspected and was new information for those who didn't believe. It showed us what our kids can do if we give them the right vehicles to do it with.

And

> When parents went through the 30-minute parent/teacher/student conferences we had this year, they focused on portfolios of student work. As a result of this kind of accountability, parents have the language and they know about standards and they know about student work. We asked parents to write reflections after the conferences. We had 240 parents write reflections . . . about the student work and conferencing. And we learned some things; they said that

> as a result of the portfolios, the report cards don't give them much information . . . This is a much higher degree of accountability. If the teacher says that this is a piece of work I value, the parent can look at it and agree and disagree. Parents are surprised . . . at what the kids can do.

Principals agreed that design-based assessments are an asset to their programs; they credited design-based tests with

- Providing information about the status and success of the program,
- Driving instruction by identifying strong and weak areas,
- Providing feedback on individual student performance and growth, and
- Reflecting the clarity (or lack thereof) of the teachers' vision of high standards.

Here are some pertinent quotes:

> We get strong information about how the NASDC program is going. It tells us that kids are working together and doing well . . . but not on specifics . . . The data don't tell us which content or skill areas we're doing well in and which we're not.

And

> The portfolios drive instruction. The teachers work and plan in clusters. They collaborate on student work . . . Weak and strong areas surface in their discussions of student work and feed into their planning.

And

> The portfolios follow students through the grades. They contain the standards and evidence to indicate completion of standards. And the evaluation of student work. They have input from parent conferences, demonstrations, video standards, and evidence and rubrics . . . The teachers score using evaluation forms. Students also keep running resumes of what they have accomplished.

Finally,

> One of the things we have done is have teachers sit in groups and look at samples of student work and define for ourselves what exemplary work is and how it was reached. The teachers talk about what led up to a particular exemplary piece. Teachers sometimes are so pleased with their own exemplars, until they see other teachers' exemplary pieces. Teachers need to have conversations with other teachers.

Reformers lamented, however, the fact that, despite the value of design-based assessment results and products, this information is not evaluated in college admissions. Again, college placement is an important, tangible accountability issue for high schools and their communities. These principals explained that most colleges do not review student portfolios in the application process. For instance,

> Purdue, Indiana University and Notre Dame said, wait a minute, we can't evaluate portfolios for all these kids. We want class rank and GPAs. They went to work politically and the state backed off. Portfolios were going to be part of IPAS. Now at the state level, portfolios officially include grades, attendance, and test scores.

SUMMARY

Principals were enthusiastic about the information obtained—about students and their programs—from design-based assessments. They said that faculty, parents, and community members get powerful information from portfolios, journals, exhibitions, conferences, and other authentic demonstrations of student learning.

Chapter Seven

RECOMMENDATIONS FOR REFORMING AND CONFORMING

ACCOUNTABILITY TESTING

Forty-five states currently have statewide assessment programs. Forty-two require administration of conventional test batteries. In 1994, 24 states offered performance-based tests. In the past year or so, several of the more prominent performance-based programs came under review by their legislatures—the Kentucky, California, Arizona, and Indiana programs are examples. Three of these programs were canceled. NASDC principals commented on the changes in state accountability testing,

> IPAS (a performance-based test under consideration in Indiana) was going to be created, but the religious right went to the legislature and quashed the test. There weren't many of them but they were vocal.

On the recently rescinded CLAS, a principal said,

> The line between CLAS and the design was direct. It made sense—that's probably why we're not doing it (administering CLAS) anymore. The district is going to have to respond (to the fact that there's not a state test). They are thinking about putting a test together that gets at the kinds of stuff CLAS looked at plus basic skills. The state has been burned and they are standing this one out.

Finally, of an existing program, a principal explained,

> If the kids perform at the level of the performance assessment given, they will be world-class learners. But the pendulum will swing, and we have to make sure that content is emphasized with process.

Several NASDC principals expressed concern that superintendents and school boards increasingly will decry performance assessments and ask for school performance information from traditional tests. Given recent legislative actions and the difficulty facing measurement professionals in developing authentic measures of the complex skills promoted by standards-based programs, their concern appears warranted. In many quarters, standardized, multiple-choice tests are seen as familiar and more easily understood than performance measures; more cost- and time-efficient; more useful because they provide individual-level scores in addition to school-level data; and possessing known psychometric properties.

ACCOUNTABILITY REPORTING

Twenty-five states currently mandate school-level profiles. Many districts require school reports as well. NASDC principals said that public reporting of accountability data through system or school report cards and by the press heightened their anxiety about the discord between accountability requirements and design tenets. They argued that many current accountability systems narrow the focus of public debate from their broad, comprehensive reform goals—and accomplishments—to things assessed by the system and profiled in the reports. The attachment of rewards and sanctions (school resources, merit pay, principal evaluations, and school take-over) to their accountability systems made principals' concerns more pointed.

RECOMMENDATIONS TO FELLOW REFORMERS

Faced with this quandary, the 20 NASDC principals were asked to offer recommendations for simultaneously reforming schools and conforming to state and district accountability systems. Their experiences from 1993 to 1995 suggested many. They offered advice to school-based reformers, jurisdiction sponsors, and directly to NASDC. Principals suggested ten strategies for restructuring in this age of accountability. Each was suggested by two or more principals.

First, they gave advice to fellow innovating principals; then they spoke to NASDC and other reform leaders. Their strategies represent differing perspectives; some are mutually exclusive. The ten recommendations are stated in the respondents' words.

Recommendations on Accountability Requirements

Six of the innovators (30 percent) offered the following advice:

- *You have to continue to take (traditional) accountability tests, but you have to look for new assessment instruments to test what you are trying to do because traditional tests do not do it justice.*

One principal advised reformers,

> The standardized tests have value, in that, in a time of change you can't throw everything out during the transition. You can't throw out the baby with the bath water. It gives a safety net to teachers and parents. You have to have two alternative systems—side-by-side. Until teachers are more confident about their work.

And

> People want to know the relativity. They need to know how to compare their child to other kids in the class, the school, the state and country. The emphasis has to be, though, on stating criteria and having assessments based on those criteria.

Although study principals spoke clearly about the limits of multiple-choice tests, most fell short of denouncing their use—at least in the absence of an operationally useful alternative. Linked to subsets of important objectives and included among other data points on school performance, respondents said traditional tests provide useful information. But, principals said, to faithfully represent their students' accomplishments, traditional test data must be coupled with information from performance assessments. The difficulties facing jurisdictions in adopting or designing tests that speak to complex, important, relevant constructs, however, are the same difficulties facing measurement professionals. Reformers and their sponsors are hampered by the nascent state of performance-based assessment.

Among the present NASDC jurisdictions, several are developing performance-based tests. Additionally, some NASDC designers are working with site partners and measurement experts to develop authentic measures that provide score links to other assessments or other testing populations. Such measures would allow principals to compare their students' performance to that of students elsewhere. Some emerging assessments also provide links to performance standards debated at the national level. If development of authentic measures is successful, the assessments will likely be enormously helpful to adopters—both in gauging school progress and fostering program improvement.

Turning to their second recommendation, seven reformers (35 percent) echoed the following suggestion:

- *If accountability data will be publicized and your school is judged by it, you have to make the school look as good as you can.*

One respondent explained,

> You don't want to short-change the school and the students in the public light. You have to put forth every effort to appear successful by whatever method you are judged by. The kids have to be as well prepared as they can. When I have to change my car tire, I don't relish it, but I'm glad I know what to do.

And

> You have to make a two-pronged attack. Even if you know that it is not right or the best thing—because of the community pieces. If you jump in and ignore them and take a hit on the community pieces, you will have a lot of trouble.

Research on the use of standardized, multiple-choice tests for accountability purposes suggests that educators can and do effectively teach to traditional tests (Shepard and Dougherty, 1991; Koretz et al., 1992; Guskey, 1994; McDonnell, 1994). By increasing instructional time on tested content and focusing on test preparation, many faculty have improved their schools' performance on conventional, high-stakes tests. Studies suggest, however, that gains may not generalize to broad improvements in students' knowledge and skills.

Recall that in response to specific questioning, 85 percent of the NASDC principals said their teachers engaged in directed preparation for state and district accountability tests; faculty prepared students for mandated assessments by reviewing content and skills likely to be on the tests, practicing test-taking skills, and working with practice tests and test preparation materials. At the close of their first year, 60 percent of the study principals reported that school data on standardized measures were favorable; none said these data represented their accomplishments at innovation. Some interviewees also talked about efforts to improve school performance on other accountability indicators (attendance rates, promotion rates, graduation rates, etc.). The modest sample size and qualitative nature of the present study do not support projections of likely outcomes for reforming schools on given accountability measures.

The success reformers will meet in "making their schools look good" on conventional tests in the absence of directed preparation is unclear. The literature on whole-school reform offers little clarity in this regard. Comprehensive studies of the performance of innovative schools on traditional tests have not been conducted. Limited information is available from two sources. In the School Restructuring Study, conducted by the Center on Organization and Restructuring Schools (CORS), test and classroom practice data for 24 restructured schools were examined (Newmann, Secada, and Wehlage, 1995). The data suggested that students in classrooms emphasizing construction of knowledge, development of in-depth understanding, and the application of learning to realistic problems performed as well as or better than comparison students on standardized tests. CORS researchers argued that it is reasonable to expect "students who think carefully about their subjects, study them in depth, and connect them to their personal experiences" to remember the facts and definitions tested by traditional measures. On the other hand, they contend, depth of coverage probably requires omission of some material normally covered in conventional classrooms. CORS researchers assert that the disadvantage should be outweighed by students' superior performance on learned material.

A review of school outcomes for selected reforms by Herman and Stringfield (1995) showed positive gains after two or three years for 80 percent of the 50 schools with whole-school programs reporting standardized test data in the literature. Further, the schools' data on

attendance, disciplinary referrals, and other accountability indices generally were positive. It is unclear, however, how these findings generalize to the reform community; it is not apparent how results reported in the literature speak to the progress of the population of over 1400 U.S. reforming schools. Whether additional reforming sites attempted to evaluate and obtained ambiguous or discouraging results and elected not to report them (the file drawer problem) is unknowable.

Given the relative infancy of authentic assessment, analogies to the data reported by CORS and in the Herman and Stringfield review have not accumulated for transforming schools on performance-based accountability tests. Even in California, Vermont, and Kentucky (the earliest states to sponsor authentic assessment programs), evidence on the impact of testing reform on student achievement is scant (Newmann and Wehlage, 1995). All three assessment efforts have experienced technical problems that threaten inferences about school improvement (Koretz et al., 1994; Cronbach, 1995; Hambleton et al., 1995).

Experts on measurement-driven instruction argue that it is much easier to narrow instruction to the content of traditional tests than to broaden it to the knowledge and skills required by performance-based tests. Practicable suggestions for maximizing school performance on state and district performance measures as recommended by the NASDC leaders are elusive. Recall that in the present study, 40 percent of the principals in states with performance-based tests reported positive results on their jurisdictions' assessments. Again, the data cannot be interpreted as outcomes for the then-rudimental programs.

Third, and on an important note, four principals (20 percent) wanted their colleagues to recognize,

- *You have to be ready for the fact that when you shake things up, there will be dips and downward trends in test and other accountability data.*

One principal said, "It's easy to see a one-year dip; it could happen if you happen to have a group with low-SES kids." The spring 1994 score data from some of these schools reinforce this message. Recall an earlier statement about CLAS results,

> In 1994 we got CLAS results and it was disappointing for all of us. We did a little better than the state and district in a couple of areas, but our math was poor . . . It wasn't what we expected. We couldn't come up with any explanations for it. We couldn't make heads or tails of it.

Another principal said, "We didn't do well this year and we are doing some soul searching about what caused our poor performance. The scores have caused us real anguish." A third respondent talked about anomalous data,

> The 1994 scores didn't make sense because we were doing high quality work and I had seen the quality of work in the classroom. We had strong teacher professional development. I knew we were doing good work. The only thing I could guess was that it had something to do with the class itself. It might have been a group that was a little weaker. I think it was an anomaly.

This issue is vexing. Many NASDC principals explained they see good things going on in their classrooms, but they have not been able to support their observations with data from the accountability tests. Needed are (1) ideas for promoting shared expectations among reform participants and school communities for the progress and outcomes of restructuring, (2) recommendations for explaining test results and their significance to the school community, and (3) plausible suggestions for responding to reform skeptics in the face of early accountability data that are lackluster or disappointing.

Fourth, respondents gave pragmatic advice for school administration in the face of conflicting expectations. They said to principals undertaking reform,

- *You should get a clear understanding of the accountability system and its correspondence to the reform program.*

One interviewee recommended,

> My advice is to have a clear understanding of the expectations of the accountability system and to make sure there is a clear match between those expectations and the restructuring program . . . make sure that they are lined up clearly. You have to make choices about what you want to do if they don't line up.

And

> If you have a personal fear of evaluation, you have to have all the parameters up front with the evaluator. It is a real stressor if you don't know what you will be accountable for.

Here, interviewees talked about dealing with jurisdiction accountability requirements head-on. They said this more vociferously in framing their recommendations to NASDC; these recommendations are described later in this report.

In direct contrast to the previous recommendations, four interviewees (15 percent) asserted that,

- *Reform takes so much energy and commitment, you can't afford to worry about accountability tests.*

A respondent explained,

> One of the things I did was not worry about it. Even if there was going to be a blip in the scores, I told myself that I'm not going to worry about it, even if the district does. Reform takes so much energy that we decided to focus everything on NASDC.

And

> The advice I would give is to let the teachers and principals know that they will not be beat over the head with test data during reform. Let them know that they are doing something different.

This stance may be workable in some jurisdictions and at some stages of reform. It allows for greater focus on design implementation. It answers interviewees who said that time diverted from reform to preparation for accountability tests is a drain on the program. This position stands in contrast to that given by the majority of principals.

Recommendations on Reforming

By way of a sixth recommendation, four principals (20 percent) gave practical advice for educating parents, community members, and

business partners about authentic assessment and the quality of students' work. They told their colleagues in reform,

- *You should bring the community into the school and get stakeholders involved in looking at student work.*

Principals made strong statements about the value of community involvement in school-based performance assessment. Principals encouraged reformers to bring in parents and others to review student portfolios, projects, and exhibitions. They said,

> Increase the importance of students' work . . . do this by getting stakeholders to talk to students about their work. When you convince stakeholders to look at student work, they are drawn in.

And

> Community members, outsiders came in and they were amazed at what the students were doing . . . The work was rich and detailed and students, teachers, and parents were surprised . . . This is an accountability system and this will help us if we get criticized for a one-year dip in MEAs.

From the principals' narratives comes the clear message that community involvement in authentic assessment is powerful. Principals encouraged reformers to involve proximate stakeholders in the review of student work. One principal said community involvement in performance-based testing was her school's "defense against" standardized testing. Respondents were emphatic in their statements about the value of this practice. Three accounts are given here.

> Parents and community members love these alternative demonstrations of student learning. We show parents videos of student learning and of the morning radio program. They tell me that they can't believe the conversations that they can have at home—because first graders and sixth graders are both getting exposed to North America . . . The parents see their kids doing and learning. And they take it as legitimate evidence of learning. Plus the projects. You have to cultivate community support. The principal has to be a PR person.

Also,

> We had our partners in education from K-Mart and Pepco and other companies come in and review exhibitions and score them. The kids loved it. They know that the teachers are paid to review their work, but when the partners came in to see their exhibitions, they knew that was important.

And

> We have a site council that includes teachers, students, parents, community members, and me that will review portfolios from the quarter. The council will make recommendations for things teachers should attend to in the future. They also will talk about the overall quality of the student work. This goes back to teachers for planning and may go into the school improvement plan. The council members review portfolios and score them on a 3-point scale . . . They scored them but what they really wanted to do was write comments on them with sticky notes.

High school principals cautioned, however, that undergraduate admissions officials may not be "drawn in" by portfolios and other authentic demonstrations of learning. Principals lamented the postsecondary communities' reluctance to accept any but conventional data in college admissions.

Seventh and on a similar note, three principals (15 percent) offered the following recommendation to their colleagues,

- *Going into reform, you should let the faculty know there will be competing forces.*

Many principals talked about their teachers "serving two masters." They reminded readers that strategies for dealing with conflicting incentives are unlikely to be available; even so, they said that providing faculty and staff with information about change and potential conflict is of value. That is,

> There should be human resource development in the school for teachers and staff . . . It should expose people to make them aware of what will happen in a time of change . . . There should be training

> on teaming, collaboration and the change process . . . Good business organizations do this.

And

> You can't stretch teachers so far without having it be non-productive. We have to let teachers know that there will be conflicting forces—but we have to make the program as strong as possible—that is the key.

RECOMMENDATIONS TO NASDC

Principals closed their interviews with suggestions to NASDC. Their recommendations also apply to the work of other reform sponsors. Principals made three recommendations. About the first they were directive; six principals (30 percent) asserted,

- *NASDC should advocate for reformers' work and ask jurisdiction officials for waivers from accountability requirements.*

Interviewees said,

> NASDC should step in and say that in order to implement these programs, you have to give the transformation a lot of attention, and the accountability system has to take a back seat.

And

> NASDC should ask the jurisdictions for flexibility for the reforming schools. They should say that if they (the systems) want the programs to work, you have to let them get going. If they (the systems) can't let go, then they're not interested in reform.

To other principals, respondents said,

> If it (standardized testing) is an optional thing, opt out. It is not complementary to the reform. It isn't a valid assessment tool for reform. Many times, that isn't an option. When you can't steer the decision—we as middle managers do what we are told while we are trying to do everything else too.

Finally,

> NASDC should issue a position statement to jurisdictions sponsoring reforms to the effect that it (standardized testing) is detrimental to the speed of the reform. We could be spending time in more progressive modes.

Respondents appealed to NASDC to advocate for their work in partner jurisdictions. Indeed, NASDC seeks to instigate system-level reform. It aspires to change operating environments in ways that promote and sustain school change. In many important arenas, NASDC has fostered the development of supportive, assistance-oriented systems. On accountability, however, they have been irresolute. Early on, they asked jurisdictions to give sites freedom to devise the means by which accountability would be demonstrated to the community; they gave the caveat that the means should go beyond, rather than supplant, those required by the state or district. But NASDC has not stepped forward to help designers and school-based reformers with the difficult task of defining meaningful indicators.

Interviewees additionally asked NASDC to negotiate on their behalf for waivers from ill-fitting jurisdiction-sponsored assessments and accountability requirements. Suggestions for doing so, however, were not forthcoming from the principals. Without introducing credible, alternative indicators of the knowledge and skills innovative programs seek to promote, the wisdom of granting waivers is questionable. Policy-makers, educators, parents, and other school stakeholders demand more and better information on the performance of schools. It is unlikely communities would long indulge reformers who do not provide information on student performance and school improvement.

In the same way that reformers must look for ways to educate proximate stakeholders about the significance of extant accountability measures, NASDC should frame statements at the system level. To date, NASDC has only generally addressed the barriers to reform posed by its partner jurisdictions' accountability systems. NASDC should give attention to developing plausible mechanisms for monitoring the progress and outcomes of restructuring. They also need to offer tools to jurisdiction sponsors for responding to reform critics if early data from system-wide measures are humbling. In a more gen-

eral sense, interviewees asked NASDC to issue and disseminate policy statements about accountability and school restructuring. The principals contend that NASDC statements would be widely read and heeded.

Second, and in a slightly different vein, respondents said that,

- *When NASDC and other observers visit transforming schools, they should provide feedback on school performance to their hosts.*

If this recommendation is followed, they said, reformers would receive valuable accountability information from fellow educators and those interested in their work. Specifically,

> It would be important for people connected with education and connected with NASDC to support people trying to be reformers . . . We have to work together . . . Good feedback should be given to schools and teachers when observers come in . . . When NASDC finds that a model is really well implemented in a school, the word should go to the district.

Indeed, one NASDC team instituted critical friends visits during school year 1994/95, providing their sites with important feedback. Also, one of the ten study jurisdictions has a critical friends component as part of their school quality review program.

And last, respondents instructed,

- *NASDC should provide us with opportunities to network with other reformers around accountability and other issues.*

One principal commented, "I want to date those schools with good practices." Another explained,

> The principals have a good overall view of our own schools, but we don't have the big picture; we don't know what other schools are doing. It would be great to be able to share with each other the kinds of things we have trouble with and share approaches. It would be wonderful to network.

Principals asked for opportunities to meet face-to-face or to hold electronic conversations with their peers. They wanted to talk about accountability and other school reform topics.

CONCLUSIONS

Rush Limbaugh (1993) argues that today's reformers do not want standards and they do not want to be held accountable. Others share his view. In *Making Schools Work* (1995), Hanushek and his colleagues talk about the aversion of many educators to fixed performance criteria and accountability. They note that several prominent programs are marked by flexible assessment systems yielding noncomparable data and by a reticence to hold educators and schools accountable.

The present study suggests, at least for the 20 NASDC principals, that these arguments are weak. These leaders do not shirk accountability; they take seriously their charge to educate students to high standards. The NASDC principals struggle with current accountability measures and assert that standardized testing works against their efforts. One principal explained,

> I wish I could look into a crystal ball and see that our kids will be better citizens and workers because of what we are doing. I would ask NASDC to be a guardian angel to the reforming schools to give them support in the community . . . But I don't know whether I can ask for that right now because this is so new and we don't know if it is the answer. I know that kids are engaged, discipline problems have dropped, attendance is up. These all say that we are moving in the right direction. Everything except for the scores.

These leaders explain that the tensions faced in simultaneously reforming and conforming to accountability systems are thick and palpable. One interviewee said,

> The accountability system has an impact on everything we do. To focus on specific basic skills, you have to drill. We would like to get away from drill and pounding stuff into kids' heads; they don't remember it the next year. But if the accountability system looks at scores to judge school effectiveness, you can't take your eyes off of basic skills. You have your feet in two countries, and you have to keep your feet in both of them.

Indeed, as the authors of *Successful School Restructuring: A Report to the Public and Educators* point out, schools are nested in a complex environment of expectations, regulations, and professional stimula-

tion from districts, state and federal agencies, universities, unions, professional organizations, the courts, and parents (Newmann and Wehlage, 1995). The report's authors state, as NASDC principals implied, that building the organizational capacity to promote high student achievement requires technical and political support from beyond the school walls. However, external influences often pull schools in different directions, impose unreasonable regulations, and instigate rapid shifts in policy—all of which can undermine organizational capacity. In the face of conflicting directives, the NASDC reformers remain hopeful about their work and committed to performance-based accountability.

More than one reform effort has been derailed by misaligned assessment or by the failure of reformers to take seriously the need to portray their progress. To date, NASDC has only generally addressed the barriers to reform posed by its partner jurisdictions' accountability systems. NASDC and others propelling innovative programs should instigate public discussions of the discord between conventional tests and teaching and learning in reforming schools. They should launch public engagement programs to establish realistic expectations about the types of accountability data that mark early implementation of whole-school designs. The campaigns should clearly recognize the legitimacy of requests for accountability data by students, teachers, parents, business partners, and the larger community.

NASDC's leaders and others propelling innovative programs should work with jurisdiction sponsors to develop methods for fairly gauging the impact of reformers' work on students and schools. Continuity with existing systems should be maintained both for the information value of extant measures and to satisfy stakeholders distrustful of rapid change and the substitution of assessments unique to reforming schools. Ill-fitting accountability requirements should be supplemented by indicators consonant with reform. Performance on a broad range of outcomes should be traced over time so that growth in achievement—for students and schools—becomes the focus. Making explicit a range of intended outcomes will thwart attempts to hold up standardized, multiple-choice scores as the hallmarks of educational success. Comprehensive progress data may help state and district sponsors as well as NASDC, Annenberg, and other reform-minded foundations and corporations (1) formu-

late policy to help transforming schools advance more quickly and effectively, and (2) present to the public the collective efforts, accomplishments, and lessons of reforming schools.

REFERENCES

Bodilly, S., *Lessons from New American Schools Development Corporation's Demonstration Phase*, RAND Institute on Education and Training, MR-729-NASDC, Santa Monica, California, 1996.

Cronbach, L., "Report of the Select Committee on the California Learning Assessment System," in *A Valedictory: Reflections on 60 Years in Educational Testing*, National Academy Press, Washington, D.C., 1995.

Guskey, T., "What You Assess May Not be What You Get," *Educational Leadership*, Vol. 51, No. 6, 1994, pp. 51–54.

Hambleton, R., R. Jaeger, D. Koretz, R. Linn, J. Millman, and S. Phillips, *Review of the Measurement Quality of the Kentucky Instructional Results Information System, 1991–1994*, Office of Education Accountability, Kentucky General Assembly, Frankfort, 1995.

Hanushek, E., et al., *Making Schools Work: Improving Performance and Controlling Costs*, Brookings Institute, Washington, D.C., 1995.

Herman, R., and S. Stringfield, *Ten Promising Programs for Educating Disadvantaged Students: Evidence of Impact*, Center for the Social Organization of Schools, Johns Hopkins University, presented at the American Educational Research Association Meeting, San Francisco, California, 1995.

Koretz, D., G. Madaus, E. Haertel, and A. Beaton, "National Educational Standards and Testing: A Response to the Recom-

mendations of the National Council on Education Standards and Testing," in *Oversight Hearings on the Report of the National Council on Educational Standards and Testing: Hearings before the Subcommittee on Elementary, Secondary, and Vocational Education of the Committee on Education and Labor*, Kentucky House of Representatives. 102d Congress, 2d Session, 1992.

Koretz, D., B. Stecher, S. Klein, and D. McCaffrey, "The Vermont Portfolio Assessment Program: Findings and Implications," *Educational Measurement: Issues and Practice*, Vol. 13, No. 3, 1994, pp. 5–16.

Limbaugh, R., *See, I Told You So*, Simon and Schuster, Inc., New York, 1993.

McDonnell, L., "Assessment Policy as Persuasion and Regulation," *American Journal of Education*, Vol. 102, No. 8, 1994, pp. 394–420.

New American Schools Development Corporation, *Designs for a New Generation of American Schools: Request for Proposals*, Washington, D.C., 1991.

New American Schools Development Corporation, *Vision to Reality: Tomorrow's Schools Today*, Washington, D.C., 1994.

New American Schools Development Corporation, *A Thousand Actions*, Washington, D.C., 1995.

Newmann, F., W. Secada, and G. Wehlage, *A Guide to Authentic Instruction and Assessment: Vision, Standards and Scoring*, Wisconsin Center for Education Research, Madison, Wisconsin, 1995.

Newmann, F., and G. Wehlage, *Successful School Restructuring: A Report to the Public and Educators by the Center on Organization and Restructuring of School*, Wisconsin Center for Education Research, Madison, Wisconsin, 1995.

Shepard, L., and K. Dougherty, "Effects of High Stakes Testing on Instruction," presented at the American Educational Research Association and the National Council on Measurement in Education Meetings, Chicago, Illinois, 1991.

A Forum for Reform

The New American Schools Development Corporation (NASDC) [illegible] the development of seven comprehensive designs for school [illegible] reform and since late 1991, these reforms have been implemented and tested in over 200 schools nationwide. NASDC asked RAND to [illegible] its impact, to examine the development, management and implementation of these school designs.

In *Reforming and Conforming: NASDC Principals Discuss School Accountability Systems*, author and researcher Susan Bodilly addresses [illegible] schools can and cannot be accountable, [illegible] accountability systems [illegible] in practice. Based on interviews with 30 of the school principals [illegible] *Reforming and Conforming* reports that the traditional standards of multiple-choice tests that are used by districts [illegible] and other measures of student performance don't address the knowledge, skills, and behaviors that innovative programs [illegible] promote.

Can school districts successfully reform while conforming to [illegible] accountability systems? How can student assessment be better [illegible] promising practices? What can NASDC do to advocate for [illegible] school reform and district policy? These and other questions [illegible] school principals, teachers, and district officials [illegible] their experiences with school reform and offer their insights [illegible] issues and more. From public meetings [illegible] to one-on-one assessment of school administrators and teachers, *Reforming and Conforming* will prove invaluable to anyone [illegible] to improve schooling.

To learn more about the whole-school designs being implemented in the NASDC test sites, see *Lessons from New American Schools Development Corporation's Demonstration Phase*, by Susan Bodilly, RAND, 1996.

ISBN 0-8330-2440-X

51500

9 780833 024404

MR-716-NASDC

60